

基于 CNN-SVR 的立体图像视觉感知评价方法

徐源, 陈纯毅*, 胡小娟, 于海洋, 田野

长春理工大学计算机科学技术学院, 吉林 长春 130022

摘要 针对立体图像的多维影响因素和预测结果准确性不足的问题, 提出一种基于卷积神经网络-支持向量回归(CNN-SVR)的立体图像视觉感知客观评价模型。该模型将基于颜色的平面显著图和基于差异的视差图相结合, 对其进行阈值分割, 得到视觉感知潜在显著不适区域; 然后进行特征提取, 分别提取对比度、颜色、结构复杂度等全局特征和视差、纹理、空间频率等局部特征; 最后采用将 CNN 和 SVR 相结合的方式构建多特征视觉感知客观评价模型, 得到最终的客观预测值。实验结果表明, 所提方法的 Pearson 相关系数高于 0.87, Spearman 相关系数高于 0.83。与现有其他方法相比, 在公开数据集上所提客观评价模型更优, 预测结果与人们主观评价结果具有更高的一致性。

关键词 立体图像; 视觉感知; 特征提取; 卷积神经网络; 支持向量回归; 客观评价模型

中图分类号 TN911.73

文献标志码 A

DOI: 10.3788/LOP230893

Visual Perception Evaluation Method of Stereo Images Based on CNN-SVR

Xu Yuan, Chen Chunyi*, Hu Xiaojuan, Yu Haiyang, Tian Ye

School of Computer Science and Technology, Changchun University of Science and Technology,
Changchun 130022, Jilin, China

Abstract This paper proposes an objective evaluation model of stereo image visual perception based on convolutional neural network (CNN) and support vector regression (SVR) to solve the issue of multidimensional influencing factors for stereo images and improve the accuracy of prediction results. In this proposed model, the plane saliency map using color and the disparity map based on differences are combined, and the potential salient discomfort regions of visual perception are obtained using threshold segmentation. Then, global features, such as contrast, color, and structural complexity, are extracted using feature extraction along with the local features, such as disparity, texture, and spatial frequency. Finally, the objective evaluation model of multifeature visual perception is constructed by combining CNN and SVR to obtain the final objective prediction value. Experimental results show that the Pearson linear correlation coefficient and Spearman's rank correlation coefficient of the proposed method are higher than 0.87 and 0.83, respectively. In addition, compared with other existing methods, the objective evaluation model proposed in this paper is better on the public dataset, and the prediction results have higher consistency with the subjective evaluation results.

Key words stereo image; visual perception; feature extraction; convolutional neural network; support vector regression; objective evaluation model

1 引言

立体图像凭借给人们带来身临其境的现场感优势得到迅速发展。立体电影、立体广告、医学影像等都通过立体图像进行展示, 但由于现有的立体图像采集方法和显示技术还不够完善, 容易给用户带来视觉不适感, 从而降低了用户的视觉感知体验^[1]。因此, 有效地模拟人类视觉系统, 对立体图像视觉感知体验进行准

确的预测和评价具有重要的研究意义和应用价值, 有助于客观准确地反映人眼在观看立体图像时的视觉感受, 同时也为优化立体图像的内容提供了帮助。

目前, 人们对立体图像的客观预测评估已经取得了一定的成果。影响立体图像视觉感知体验的因素有很多, 例如双目视差、辐辏调节冲突、左右图像不对称、心理因素及观看条件等^[2-3]。根据国内外的研究现状, 本文将从双目视差信息、左右视图的特征、结合图像显

收稿日期: 2023-03-19; 修回日期: 2023-03-21; 录用日期: 2023-03-23; 网络首发日期: 2023-03-30

基金项目: 国家自然科学基金项目(U19A2063)、吉林省科技发展计划项目(20230201080GX)

通信作者: *chenchunyi@hotmail.com

著性及深度学习的角度来展开描述。

首先人眼的双目视差是影响立体视觉感知体验最主要的原因^[4]。胡佳洁等^[5]为了探究视差因素对立体图像视觉感知体验程度的影响,进行了大量的主观实验,得到了立体图像舒适的视差范围。其研究成果具有实用价值,为立体图像的特征提取提供了有效的依据。

为进一步提高客观评价模型的预测性能,研究者们开始结合左右视图特征进行进一步的研究。李一凡等^[6]通过四元数小波变换对左右视图进行分解,并结合人眼的视觉感知特性生成单目图,从单目图中提取感知特征,最后通过 XGBoost 回归模型进行预测。此方法针对双目视图信息进行客观评价并取得了较好的结果,但是他们只关注了双目视图之间的联系与区别导致视觉体验下降的因素,并没有从立体图像视觉感知的多维影响因素进行考虑。

考虑到人眼在观看立体图像时会更加关注图像中最显著的区域,因此研究者们便开始结合显著性进行研究。Zhang 等^[7]提取了显著性加权视差特征,并利用对深度图和平面显著图进行自适应线性融合的方式,得到视觉注意模型。但是简单的线性组合会改变原始的视差范围,同时也会减弱图像中显著信息的效果,因此所提模型通过将平面显著图作为视差图的权重,得到立体视觉显著图,这样可以更加有效地保持原有的显著信息和视差信息。

近些年,随着深度学习在图像处理领域取得了良好的成果,研究者们把目光转至利用神经网络来进行预测的方面。Kim 等^[8]提出了基于双目融合深度网络的客观评价方法,将方法用于学习潜在的双目特征,并设计了视差正则化网络来对预测结果进行改善。然而仅采用深度学习网络进行回归预测得到的模型性能不

如支持向量回归(SVR)强大。因此本文采用将卷积神经网络(CNN)和 SVR 相结合的方式构建立体图像视觉感知客观评价模型。

综上所述,本文针对立体图像视觉感知的多维影响因素和预测准确性问题,提出了一种新的立体图像视觉感知客观评价模型。采用将立体图像的显著信息和视差信息相结合的方式,得到视觉感知潜在显著不适区域,并提取该区域的局部特征和立体图像的全局特征,该方法不仅反映了立体图像的视觉显著性和视觉舒适的程度,同时考虑到了立体图像视觉感知的多维影响因素。采用 CNN 和 SVR 相结合的方式构建立体图像客观评价模型,该模型充分考虑到不同类型特征之间的联系,并且预测结果准确性较高。

2 基于 CNN-SVR 的立体图像视觉感知客观评价模型

提出的 CNN 和 SVR 相结合的立体图像视觉感知客观评价模型首先考虑到人眼视觉注意机制和双目融合的极限,结合了基于颜色的平面显著图和基于差异的视差图;再通过阈值分割方法得到视觉感知潜在显著不适区域,该区域在很大程度上决定了立体图像的整体视觉感知舒适度;其次在特征提取的过程中,考虑到影响立体图像视觉感知的诸多影响因素,提取了立体图像的多维特征,分别是基于图像的全局特征和局部特征,其中基于图像的全局特征包括图像对比度、颜色特征和结构复杂度,基于图像的局部特征包括视差、纹理、空间频率;最后,利用 CNN-SVR 模型对得到的所有特征进行融合,构建立体图像客观评价模型,通过预测函数计算出客观预测值。客观评价模型的总体框架如图 1 所示。

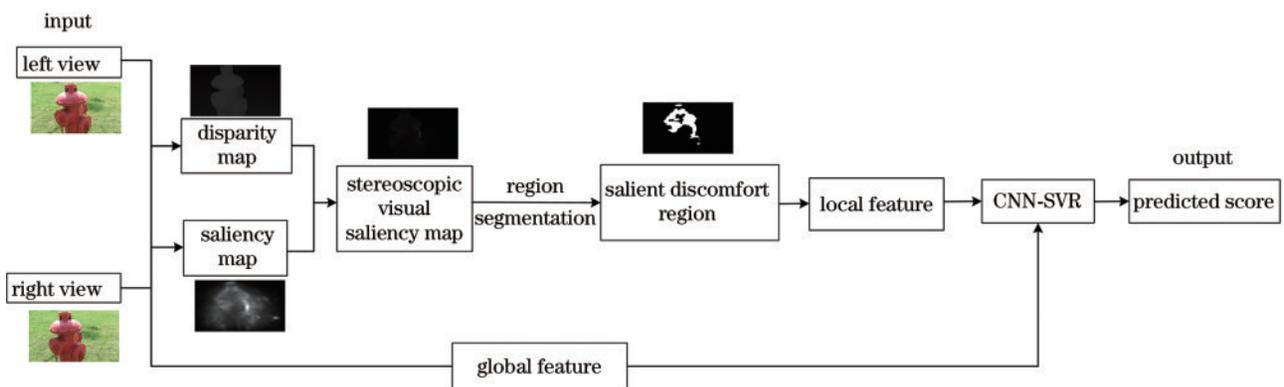


图 1 客观评价模型的总体框架

Fig. 1 Overall framework of the objective evaluation model

2.1 视觉感知潜在显著不适区域

考虑到人眼的视觉注意机制,人们在观看一幅图像时,人类的视觉系统(HVS)总是会将视线聚焦于图像中最显著的区域,所以所提客观评价模型采用图像视觉显著性(GBVS)算法^[9]提取基于颜色的平

面显著图 S_s 。同时,视觉感知舒适程度大多与基于差异的显著性有关。过度的差异会给观看者带来令人讨厌的体验,也就是说,立体图像中包含的视差越大,观看者感觉越不舒服,所以所提客观评价模型用光流算法^[10]提取基于差异的视差图 S_d 。然后将基于

颜色的平面显著图和基于差异的视差图相结合,得到立体视觉显著图。这里,将 S_s 作为 S_d 的权重,而不是它们的线性组合来获得立体视觉显著图。因为对平面显著图和视差图进行简单的线性组合会减弱图像中颜色显著性信息的效果,在很大程度上改变了图像中原始视差的范围。立体视觉显著图 S_{vs} 的表达式^[11]为

$$S_{vs}(x, y) = S_s(x, y) * S_d(x, y) \quad (1)$$

视觉感知潜在显著不适区域由立体视觉显著图 S_{vs} 进行区域分割得到,其可以有效地反映视觉感知舒

适的程度和视觉显著性。选择立体视觉显著图 S_{vs} 中最高 10% 的值作为显著不适区域的阈值,其中 T 为阈值。显著不适区域 R 的定义^[11]为

$$R = \begin{cases} 1, & |S_{vs}| > T \\ 0, & |S_{vs}| \leq T \end{cases} \quad (2)$$

上述处理方法得到的视觉感知潜在显著不适区域的获取过程如图 2 所示,其中图 2(a)和图 2(b)分别为立体图像的左视图和右视图,图 2(c)~(f)分别对应立体图像的视差图、平面显著图、立体视觉显著图、显著不适区域。

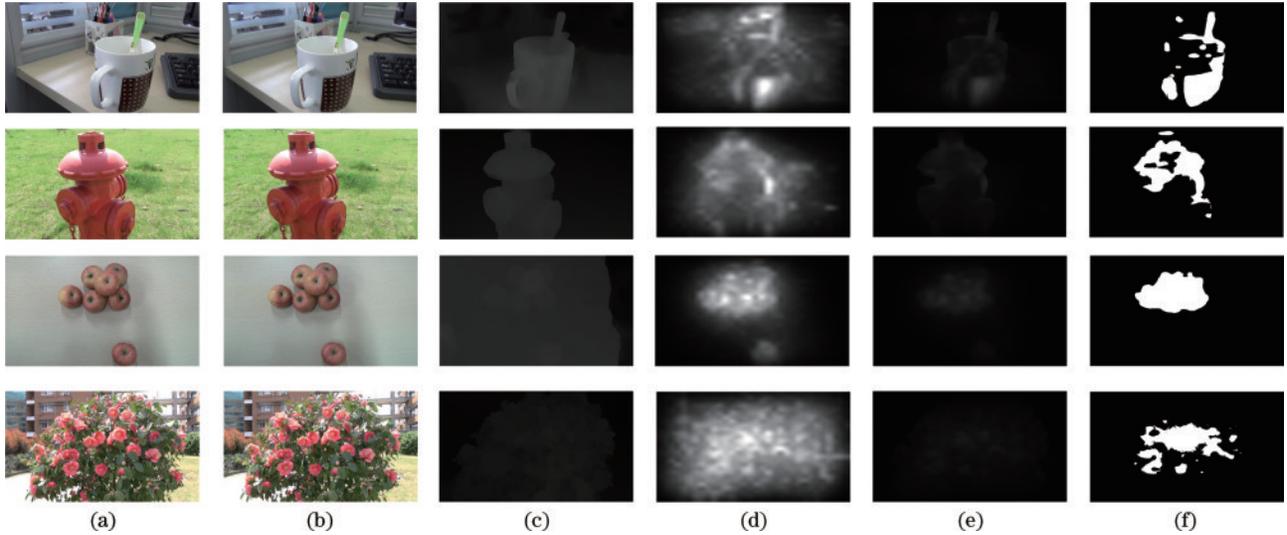


图 2 潜在显著不适区域的获取过程。(a)左视图;(b)右视图;(c)视差图;(d)平面显著图;(e)立体视觉显著图;(f)显著不适区域
Fig. 2 Acquisition process of potential significant discomfort area. (a) Left view; (b) right view; (c) disparity map; (d) planar saliency map; (e) stereoscopic saliency map; (f) salient discomfort region

2.2 全局特征

在特征提取的过程中,特别考虑了影响立体图像视觉感知体验的多维因素,分别提取基于图像的全局特征和局部特征。其中立体图像的全局特征分为 3 部分,包括对比度、颜色特征及结构复杂度。

2.2.1 对比度

对比度指图像中最亮区域和最暗区域之间的不同亮度等级的测量,一般用灰度等级的程度来衡量。在立体图像中,对比度会给立体图像的深度感知带来很大的变化,并且能够影响用户观看立体图像时视觉感知的程度^[12]。所以,对于不同程度的对比度,立体图像的主观评分值也会有很大的差异。因此,对比度是影响人们观看立体图像感知体验的一个影响因素,采用的对比度计算公式^[13]为

$$C = \sum_i^m \sum_j^n \left\{ \delta[(i, j), (i', j')]^2 P[(i, j), (i', j')] \right\} \quad (3)$$

式中: (i', j') 表示围绕中心像素 (i, j) 最近的 8 个像素; $P(i, j)$ 是每个像素点的灰度值; $\delta[(i, j), (i', j')] =$

$P(i, j) - P(i', j')$ 表示立体图像对中右视图的每个中心像素点与其周围 8 个像素点之间的灰度差; C 表示图像对比度; m 和 n 分别表示矩阵的行和列。

2.2.2 颜色特征

人类的视觉系统对颜色信息是极其敏感的。当立体图像的左右视图中颜色信息无法准确同步时,就会引起双目竞争,人眼会很难产生立体感,从而造成视觉感知体验下降和不适感。

HSV 颜色空间中包含了三个颜色分量,分别是色调 (hue)、饱和度 (saturation) 和亮度 (value)。由于 HSV 颜色空间中各个分量之间相互独立,更符合人类视觉系统的感知特性,而 RGB 颜色空间的各分量之间具有较高的相关性^[14]。因此本文利用 HSV 颜色空间提取颜色特征,并将其作为立体图像的感知特征。首先将立体图像由 RGB 颜色空间转换到 HSV 颜色空间,计算公式^[14]为

$$V = \max(R, G, B) \quad (4)$$

$$H = \begin{cases} \frac{60 \times (G - B)}{V - \min(R, G, B)}, & V = R \\ 120 + \frac{60 \times (B - R)}{V - \min(R, G, B)}, & V = G \\ 240 + \frac{60 \times (R - G)}{V - \min(R, G, B)}, & V = B \\ 360 + H, & H < 0 \\ 0, & V = \min(R, G, B) \end{cases}, \quad (5)$$

$$S = \begin{cases} \frac{V - \min(R, G, B)}{V}, & V \neq 0 \\ 0, & V = 0 \end{cases}. \quad (6)$$

从 HSV 颜色空间图像中分别提取亮度特征 F_v 、饱和度特征 F_s 及色调特征 F_h , 计算公式分别为

$$F_v = \frac{1}{N} \sum_{i=0}^a \sum_{j=0}^b V(i, j), \quad (7)$$

$$F_s = \frac{1}{N} \sum_{i=0}^a \sum_{j=0}^b S(i, j), \quad (8)$$

$$F_h = \frac{1}{N} \sum_{i=0}^a \sum_{j=0}^b H(i, j). \quad (9)$$

2.2.3 结构复杂度

当人们在观看结构较为复杂的立体图像时, 会由于图像中关注物过多, 难以形成立体感, 引起视觉不适, 造成视觉感知体验下降。这是由于图像的低频信息决定了图像的基本结构, 因此本文先利用小波变换得到立体图像的低频信息, 再对图像的低频信息进行重构。即进行小波变换后利用系数的均值进行重构, 将小波系数中小于均值的设置为 0, 接着对保留下来的小波系数进行重构, 进而得到了只包含低频信息的立体图像。结构复杂度是通过比较重构后的图像与原始图像之间的关系得来的, 计算公式^[15]为

$$F = \frac{\sum_i \sum_j (I_{mg} - I'_{mg})^2}{\sum_i \sum_j (I'_{mg})^2}, \quad (10)$$

式中: I_{mg} 表示原始立体图像, I'_{mg} 表示重构后的立体图像。重构后的部分图像如图 3 所示。



图 3 重构后的部分图像

Fig. 3 Part of the reconstructed images

2.3 局部特征

提取的立体图像的局部特征也分为 3 部分, 包括视差、纹理及空间频率特征。

2.3.1 视差特征

人们在观看立体图像时, 双眼焦点的调节与辐辏冲突是引起视觉不适的主要原因^[16]。这是由于人们在观看立体影像时, 人眼的成像点与屏幕真实的成像点之间存在差异, 难以形成立体感, 从而带来视觉不适等现象。通过调研文献发现, 人们在主观视觉感知体验评估的研究中已经报告了人眼的视觉不适会随着视差的增加而增加的内容^[17]。所以, 视差在立体视觉中起着重要作用, 是影响用户视觉感知体验的关键因素。因此在视觉感知潜在显著不适区域的基础上结合视差图, 生成潜在显著不适区域视差图, 并在该区域提取视差特征, 部分潜在显著不适区域视差图如图 4 所示。

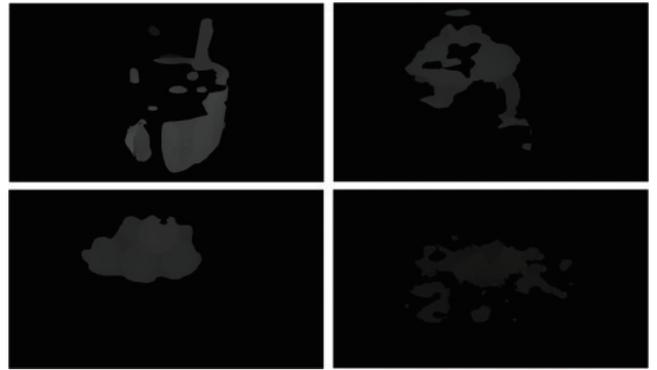


图 4 潜在显著不适区域视差图

Fig. 4 Disparity map of potential significant discomfort area

在潜在显著不适区域视差图中提取视差均值 γ ^[18]、视差角 θ ^[16] 作为感知特征, 计算公式分别为

$$\gamma = \frac{1}{N} \sum_{(x,y) \in R} S_d(x, y) * S_{vs}(x, y), \quad (11)$$

$$\theta = 2 \arctan \frac{ky}{2L}, \quad (12)$$

式中: N 表示像素数量; k 为立体图像的放映放大率; L 为用户观看立体图像时与屏幕的距离。

2.3.2 纹理特征

当立体图像中存在过于复杂的纹理信息时, 立体图像中剧烈变化的区域增多。人们在观看这种纹理较为复杂的立体图像时, 左右眼会难以形成立体感, 进而造成视觉感知不适。因此纹理特征能够影响用户观看图像时的视觉感知舒适度。采用灰度共生矩阵方法^[19], 分别提取了能量 (E_{ASM})、熵 (V_{ENT})、惯性矩 (V_{con})、标准方差 (V_{var}) 4 种能够代表立体图像纹理信息的特征作为立体图像预测评估模型的感知特征, 计算公式^[20]分别为

$$E_{ASM} = \sum_i \sum_j P(i, j)^2, \quad (13)$$

$$V_{\text{ENT}} = - \sum_i \sum_j P(i, j) \log [P(i, j)], \quad (14)$$

$$V_{\text{con}} = \sum_i \sum_j (i - j)^2 P(i, j), \quad (15)$$

$$V_{\text{var}} = \sqrt{\sum_i \sum_j P(i, j) (i - V_{\text{mean}})^2}, \quad (16)$$

式中: $P(i, j)$ 表示灰度共生矩阵; V_{mean} 表示立体图像的平均灰度值。

2.3.3 空间频率

通过调研文献发现, 立体图像的空间频率能够影响人眼双目融合的极限, 从而影响人们观看立体图像时的视觉感知体验。具有低空间频率的立体图像所产生的模糊现象会导致人类视觉系统的融合范围增加。因此具有高空间频率的立体图像会比具有低空间频率的立体图像更能导致视觉感知的不舒适, 影响人们的观看体验^[21]。所提取的立体图像空间频率图 F 和均值

η 的计算公式^[22]分别为

$$|F| = \sqrt{H(x, y)^2 + V(x, y)^2 + D(x, y)^2}, \quad (17)$$

$$\eta = \frac{1}{N} \sum_{(x, y) \in R} F(x, y) * S_{\text{vs}}(x, y), \quad (18)$$

式中: $H(x, y)$ 、 $V(x, y)$ 和 $D(x, y)$ 分别表示水平、垂直和对角方向的频率。

2.4 回归模型构建

在之前的研究中, 已经获得了立体图像视觉感知的多维特征, 现对这些多维特征进行融合, 构建视觉感知客观评价模型, 进而得到模型的最终预测结果。在立体视觉预测中, 大多数研究者会利用非线性回归方式进行融合预测, 因为非线性回归融合相比传统的线性回归融合能更有效地将所得到的多维感知特征组合起来。本文采用将 CNN^[23] 和 SVR^[24] 相结合的方式构建回归预测模型。CNN-SVR 的回归预测过程如图 5 所示。

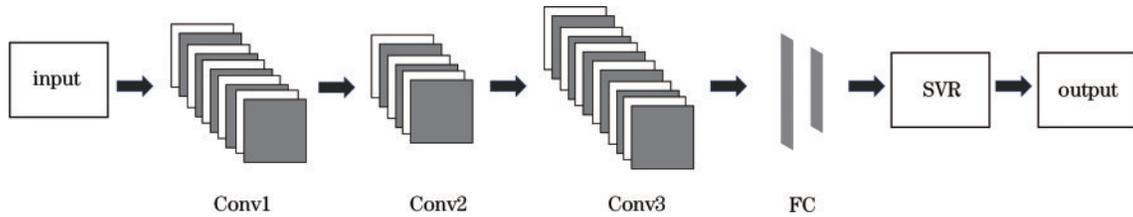


图 5 CNN-SVR 的回归预测过程

Fig. 5 Regression prediction process of CNN-SVR

所构建的 CNN-SVR 模型将获得的多维度视觉感知特征作为 CNN 的输入样本, 在将特征输入到 CNN 之前, 首先要对所提取的特征进行预处理, 这里采用的预处理方式是将所提取的一维感知特征向量以龙摆尾形式排列成特征矩阵:

$$X = [x_1, x_2, x_3, \dots, x_n] = \begin{bmatrix} x_1 & \vdots & \dots & \vdots \\ x_2 & \vdots & \dots & \vdots \\ \vdots & x_{s+1} & \dots & x_{n-1} \\ x_{s-1} & x_s & \dots & x_n \end{bmatrix}, \quad (19)$$

式中: n 为所提取的特征个数; s 为所提取的第 s 个特征数据。该方法更加综合地考虑了全局和局部两个维度的特征对立体图像视觉感知的影响, 有效地将所提取的各类视觉感知特征联系起来, 充分考虑了不同类型特征之间的联系。

将处理后的特征矩阵作为 CNN 的输入样本, 其中 CNN 主要包括 3 个卷积层和 1 个全连接层。设置的 3 个卷积层中的卷积核个数分别为 16、20 和 32, 大小为 3×2 , 卷积的步长为 1×1 , 激活函数为 ReLU。在训练的过程中, 首先将数据集做打乱处理, 数据集中 80% 的数据用于训练, 20% 的数据用于测试, 并对数据进行归一化。每次训练样本的个数为 30, 最大训练次数为 500, 初始学习率为 0.01, 并且训练 400 次后学习率下降, 下降因子为 0.1, 因此学习率变为 0.01×0.1 。

最后将卷积神经网络的输出向量作为 SVR 的输入样本, 进而对立体图像的视觉感知分数进行客观预测。在 SVR 模型中选择径向基函数作为核函数, 并通过网格搜索 (grid search) 来寻找最优参数^[25], 利用 Scikitlearn 包里面的网格搜索 GridSearchCV 函数, 以寻找最优参数; 对各个参数可能的取值进行排列组合, 列出所有可能的组合结果, 然后将各个组合用于 SVR 训练, 自动调整至最佳参数组合, 最后确定并使用该组参数。本文参数中 $C=64, \gamma=1, \epsilon=0.1$ 。

3 实验结果分析

3.1 数据集

为验证所提模型的客观性能, 采用的数据集为 NBU S3D-VCA^[26] 和 IVY LAB^[27] 立体图像数据库。数据库分别包含 200 对和 120 对立体图像及对应的平均主观评分值 (MOS)。MOS 越高, 代表立体图像的视觉感知舒适程度越高。两个立体图像数据库都包含了室内场景和室外场景, 并且每幅立体图像的分辨率均为 1920×1080 。图 6 展示了两个公开立体图像数据库中的部分图像。

3.2 性能指标

在模型性能评估方面, 通过计算客观评价模型的预测值与主观真实值之间的相关性来衡量客观评价模型的性能。共包括 3 组客观参量, 分别为 Pearson 相关系数 (PLCC)、Spearman 相关系数 (SROCC) 及均方根



图 6 公开立体图像数据库中的部分图像

Fig. 6 Part images in the public stereo image databases

误差(RMSE)。三者可以准确地表示模型客观预测值与主观真实值之间的相关性和误差。其中, Pearson相关系数和 Spearman 相关系数用于衡量模型客观预测值和主观真实值之间的线性关系和单调关系, 均方根误差用于衡量模型客观预测值的准确度。Pearson 相关系数和 Spearman 相关系数越接近于 1, 均方根误差越接近于 0, 表明客观评价模型的性能越好。

3.3 性能比较

为了验证所提模型的性能, 对所提模型与近几年先进的客观评价模型分别在 NBU S3D-VCA 和 IVY LAB 立体图像数据库上进行性能比较。具体的实验结果如表 1 和表 2 所示。分别计算出所提模型的客观

表 1 NBU S3D-VCA 数据库中不同模型性能对比

Table 1 Performance comparison of different models on the NBU S3D-VCA dataset

Model	PLCC	SROCC	RMSE
Model in Ref. [8]	0.8128	0.7684	0.4015
Model in Ref. [28]	0.8240	0.7870	0.5580
Model in Ref. [2]	0.8310	0.7980	0.2780
Model in Ref. [29]	0.8358	0.8131	0.4973
Model in Ref. [3]	0.8668	0.8034	0.4732
Proposed model	0.8759	0.8335	0.4010

表 2 IVY LAB 数据库中不同模型性能对比

Table 2 Performance comparison of different models on the IVY LAB dataset

Model	PLCC	SROCC	RMSE
Model in Ref. [30]	0.7361	0.7190	0.4704
Model in Ref. [2]	0.8290	0.8300	0.2350
Model in Ref. [31]	0.8358	0.8131	0.4973
Model in Ref. [3]	0.9315	0.8976	0.2641
Proposed model	0.9469	0.9035	0.2547

预测结果与主观真实评价结果的 PLCC、SROCC 和 RMSE, 并与近几年先进的客观评价模型进行对比。从表 1 和表 2 可知: 所提模型的 PLCC、SROCC、RMSE 三个性能指标在两个公开数据集都得到了较好的结果, 其中, PLCC 值高于 0.87, SROCC 高于 0.83, RMSE 低于 0.41。

Kim 等^[8]利用双目融合深度网络提取双目感知特征, 进而预测立体图像的视觉舒适分数。Zhou 等^[28]引入混合失真, 利用 SVR 对所提取的特征进行融合, 得到客观预测值。Karimi 等^[2]通过聚合双目视差图提取特征, 并采用稀疏特征的无监督学习方法, 通过对 SVR 模型进行训练和测试得到客观预测值。虽然所提方法的 RMSE 指标不如 Karimi 等^[2]的方法, 但是 PLCC 和 SROCC 两个性能指标优于 Karimi 等^[2]的方法, 并且在特征提取上考虑得更为全面, 提取了更多影响立体图像的视觉感知特征。Zhou 等^[29]采用多尺度离散余弦变换(DCT)系数, 从立体图像的视差图中提取影响视觉舒适度的关键特征, 并采用随机森林回归模型进行预测。Yang 等^[3]利用三种角视差图, 分别提取了像素视差、角度视差和神经活动特征, 并采用两个 SVR 模型进行视觉感知预测。Jiang 等^[30]提取视差偏度、视差大小、视差离散度和视差对比度等作为视觉感知特征, 最后利用最大池化来预测视觉不适评分。Xu 等^[31]通过提取视差和空间频率等特征, 以视觉不适的主观值为约束条件, 构建监督字典学习算法, 进而获得客观视觉感知评分值。

以上几种先进的客观评价模型的视觉感知评分均是由单个回归预测模型进行特征融合得到的。为了全面比较所提方法与先进模型的性能, 所提方法提取立体图像的全局特征和局部特征两大类特征, 并且除了使用 SVR 模型, 还使用 CNN 对所提取的特征进行融合, 再将其输入到 SVR 模型中进行预测, 得到的预测

结果较优。综上所述,所提模型在性能上较优,并且客观预测值与主观真实值之间的误差较小,能够更加真实地反映出人们观看立体图像时视觉感知的体验程度。

图 7 给出了所提模型在 NBU S3D-VCA 数据库上得到的立体图像的客观预测值与对应的主观评分值之间的关系散点图。从图 7 可以清晰地看出,散点分布较为集中,说明所提客观评价模型的预测分数与主观评分值密切相关,具有较高的相关性。并且预测结果较为准确,证明所提模型表现出了很好的性能,能够较好地评估立体图像的视觉感知体验程度。

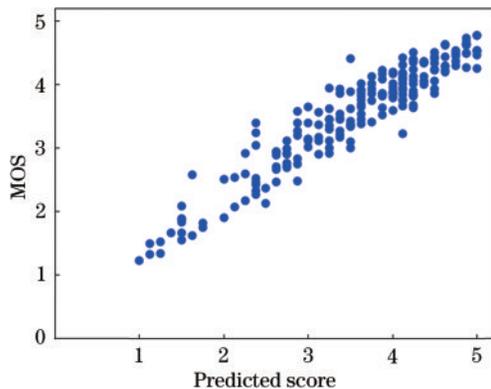


图 7 模型预测值和主观评分值的关系散点图

Fig. 7 Scatter plot of the relationship between model-predicted value and subjective evaluation value

3.4 回归方式比较

为了验证所提模型运用 CNN-SVR 作为回归方式在预测过程中的优越性,分别采用 CNN、SVR 及 CNN-SVR 作为回归方式,其中 CNN 回归预测模型的结构参数与 CNN-SVR 大体保持一致,不同的是在全连接层之后接入 dropout 层,并且设置丢弃率为 0.2,最后再接入回归层进行预测,选择 Linear 为激活函数。SVR 预测模型的核函数同样为径向基函数,并且同样采用网格搜索的方式寻找模型的最优参数。在 NBU S3D-VCA 和 IVY LAB 两个公开数据库中进行实验,对得到的结果进行比较,获得的实验结果如表 3 所示。

表 3 不同回归方式的性能对比

Table 3 Performance comparison of different regression methods

Regression method	NBU S3D-VCA			IVY LAB		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
CNN	0.7886	0.7284	0.5973	0.8336	0.8037	0.4375
SVR	0.8617	0.8162	0.5428	0.8945	0.8672	0.3560
CNN-SVR	0.8759	0.8335	0.4010	0.9469	0.9035	0.2547

从表 3 可以清晰地看出,在 NBU S3D-VCA 和 IVY LAB 数据库中,相比独立的 CNN 回归模型和 SVR 模型,CNN-SVR 模型的各个性能指标最优。因此采用 CNN-SVR 作为所提模型的回归方法。

3.5 不同类型特征之间的对比

所提立体图像视觉感知客观评价模型在特征提取的过程中主要提取了图像的两大类特征,分别是全局特征和局部特征。为了进一步展示特征提取方法的有效性,在 NBU S3D-VCA 数据集上分别对比只提取局部特征的方法、只提取全局特征的方法、两特征相结合的提取方式。实验对比结果如表 4 所示。

表 4 不同特征提取方法的对比

Table 4 Comparison of different feature extraction methods

Method	PLCC	SROCC	RMSE
Global feature	0.7271	0.6307	0.5994
Local feature	0.8173	0.7614	0.5042
Proposed method	0.8759	0.8335	0.4010

从表 4 可以清晰地看出,仅提取立体图像的局部特征或全局特征时,客观评价模型的性能下降,模型不能达到最佳的性能。这说明所提取的全局特征和局部特征对模型都起着关键性作用,两者相辅相成。因此将两者结合作为所提模型的感知特征是非常合理并且准确的,有效地提高了立体图像客观评价模型的预测准确性,进而证明所提模型的有效性。

4 结 论

提出了一种基于 CNN-SVR 的立体图像视觉感知客观评价模型。该模型有效地考虑了影响立体图像视觉感知的多维影响因素,将立体图像的平面显著图和视差图相结合,得到视觉感知潜在显著不适区域,再分别提取立体图像的全局特征和局部特征。利用 CNN-SVR 对所提取的多维特征进行融合,回归预测得到立体图像的客观预测值。在两个公开数据集上对所提模型与近几年先进的客观评价模型进行实验分析与对比,所提模型的 RMSE 值最多减少了 28.14%,说明所提模型的预测准确性更高。使用 CNN-SVR 组合模型作为回归方式得到的结果相较于其他回归方式更优,PLCC 相比 CNN 回归模型提升了 13.6%。因此所提客观评价模型性能较优,并且能够更好地反映人眼的主观视觉感受。

所提视觉感知客观评价模型采用的数据集均为原始立体图像,不包含各种模拟的失真立体图像,目前还不能说明所提模型对失真立体图像的有效性,后续将会在失真立体图像数据集上进行实验。所提模型重点考虑了人眼的视觉感知特性,没有考虑到神经特征和心理立体视觉特征等其他因素对立体图像视觉感知体验的影响。后续实验将考虑在模型中加入更多可能影响视觉感知体验的因素,以进一步提高模型预测的准确性。

参 考 文 献

- [1] Jung C, Cao L H, Liu H M, et al. Visual comfort enhancement in stereoscopic 3D images using saliency-adaptive nonlinear disparity mapping[J]. *Displays*, 2015, 40: 17-23.
- [2] Karimi M, Nejati M, Lin W S. Bi-disparity sparse feature learning for 3D visual discomfort prediction[J]. *Signal Processing*, 2021, 188: 108179.
- [3] Yang J C, Nguyen V, Sim K, et al. 3-D visual discomfort assessment considering optical and neural attention models[J]. *IEEE Transactions on Broadcasting*, 2020, 66(2): 279-291.
- [4] Oh H, Lee S, Bovis A C. Stereoscopic 3D visual discomfort prediction: dynamic accommodation and vergence interaction model[J]. *IEEE Transactions on Image Processing*, 2016, 25(2): 615-629.
- [5] 胡佳洁, 李素梅, 常永莉, 等. 基于显著区域的立体图像舒适视差范围研究[J]. *光学学报*, 2018, 38(8): 0811001.
Hu J J, Li S M, Chang Y L, et al. Comfortable disparity range of stereo images based on the salient region[J]. *Acta Optica Sinica*, 2018, 38(8): 0811001.
- [6] 李一凡, 李朝锋, 桑庆兵. 四元数小波变换优化单目图的无参考立体图像质量评价[J]. *激光与光电子学进展*, 2019, 56(18): 181006.
Li Y F, Li C F, Sang Q B. No-reference stereo image quality assessment of cyclopean images optimized using the quaternion wavelet transform[J]. *Laser & Optoelectronics Progress*, 2019, 56(18): 181006.
- [7] Zhang W L, Luo T, Jiang G Y, et al. Using saliency-weighted disparity statistics for objective visual comfort assessment of stereoscopic images[J]. *3D Research*, 2016, 7(2): 17.
- [8] Kim H G, Jeong H, Lim H T, et al. Binocular fusion net: deep learning visual comfort assessment for stereoscopic 3D[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(4): 956-967.
- [9] Harel J, Koch C, Perona P. Graph-based visual saliency [C]//*Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, December 4-7, 2006, Vancouver, British Columbia, Canada. Cambridge: MIT Press, 2006, 19: 545-552.
- [10] Sun D Q, Roth S, Black M J. Secrets of optical flow estimation and their principles[C]//*2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 2432-2439.
- [11] Jung C, Liu H M, Cui Y. Visual comfort assessment for stereoscopic 3D images based on salient discomfort regions[C]//*2015 IEEE International Conference on Image Processing (ICIP)*, September 27-30, 2015, Quebec City, QC, Canada. New York: IEEE Press, 2015: 4047-4051.
- [12] Ren H, Su Z B, Lü C H, et al. The effect of region contrast on the visual comfort of stereoscopic images[J]. *Electronics Letters*, 2015, 51(13): 983-985.
- [13] Su Z B, Li D R, Liu B, et al. A visual comfort assessment approach of stereoscopic images based on a random forest regressor[C]//*2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, June 12-14, 2020, Chongqing, China. New York: IEEE Press, 2020: 1456-1461.
- [14] 常化文, 陈春香. 基于 HSV 变换与小波变换的遥感图像融合[J]. *计算机工程与设计*, 2007, 28(23): 5682-5684.
Chang H W, Chen C X. Image fusion based on the HSV color space model and wavelet transform[J]. *Computer Engineering and Design*, 2007, 28(23): 5682-5684.
- [15] 梁化青, 周洋, 崔金鹏, 等. 一种新型立体图像舒适度预测方法: CN111696076A[P]. 2020-09-22.
Liang W Q, Zhou Y, Cui J P, et al. The novel three-dimensional image comfort prediction method: CN111696076A [P]. 2020-09-22.
- [16] Lee S I, Jung Y J, Sohn H, et al. The effect of stimulus width on the perceived visual discomfort in viewing stereoscopic 3-D-TV[J]. *IEEE Transactions on Broadcasting*, 2013, 59(4): 580-590.
- [17] Kim T, Kang J, Lee S, et al. Multimodal interactive continuous scoring of subjective 3D video quality of experience[J]. *IEEE Transactions on Multimedia*, 2014, 16(2): 387-402.
- [18] 姜求平, 邵枫, 蒋刚毅, 等. 基于视觉重要区域的立体图像视觉舒适度客观评价方法[J]. *电子与信息学报*, 2014, 36(4): 875-881.
Jiang Q P, Shao F, Jiang G Y, et al. An objective stereoscopic image visual comfort assessment metric based on visual important regions[J]. *Journal of Electronics & Information Technology*, 2014, 36(4): 875-881.
- [19] 刘丽, 匡纲要. 图像纹理特征提取方法综述[J]. *中国图象图形学报*, 2009, 14(4): 622-635.
Liu L, Kuang G Y. Overview of image textural feature extraction methods[J]. *Journal of Image and Graphics*, 2009, 14(4): 622-635.
- [20] Haralick R M, Shanmugam K, Dinstein I. Textural features for image classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3(6): 610-621.
- [21] Kim D, Sohn K. The visual fatigue prediction for stereoscopic image[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(2): 231-236.
- [22] 马恒, 贾传炎, 刘爽. 基于空间频率的图像融合方法[J]. *计算机工程与应用*, 2005, 41(31): 215-218.
Ma H, Jia C Y, Liu S. Image fusion based on spatial frequency[J]. *Computer Engineering and Applications*, 2005, 41(31): 215-218.
- [23] Jin K H, McCann M T, Froustey E, et al. Deep convolutional neural networks for inverse problems in imaging[J]. *IEEE Transactions on Image Processing*, 2017, 26(9): 4509-4522.
- [24] Liu T J, Liu K H, Shen K H. Learning-based no-reference metric for assessing the quality of experience of

- stereoscopic images[J]. *Journal of Visual Communication and Image Representation*, 2019, 61: 272-283.
- [25] 何彦, 肖圳, 李育锋, 等. 使用 CNN-SVR 的汽车组合仪表组装质量预测方法[J]. *中国机械工程*, 2022, 33(7): 825-833.
- He Y, Xiao Z, Li Y F, et al. An assembly quality prediction method for automotive instrument clusters using CNN-SVR[J]. *China Mechanical Engineering*, 2022, 33(7): 825-833.
- [26] Jiang Q P, Shao F, Jiang G Y, et al. Three-dimensional visual comfort assessment via preference learning[J]. *Journal of Electronic Imaging*, 2015, 24(4): 043002.
- [27] Jung Y J, Sohn H, Lee S I, et al. Predicting visual discomfort of stereoscopic images using a human attention model[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, 23(12): 2077-2082.
- [28] Zhou Y, Chen Z B, Li W P. Hierarchical visual comfort assessment for stereoscopic image retargeting[J]. *Signal Processing: Image Communication*, 2021, 95(3): 116236.
- [29] Zhou Y, Yu W L, Li Z, et al. A stereoscopic visual discomfort prediction using multi-scale DCT features [C]//*Proceedings of the 27th ACM International Conference on Multimedia*, October 21-25, 2019, Nice, France. New York: ACM Press, 2019: 184-191.
- [30] Jiang Q P, Shao F, Jiang G Y, et al. Visual comfort assessment for stereoscopic images based on sparse coding with multi-scale dictionaries[J]. *Neurocomputing*, 2017, 252: 77-86.
- [31] Xu H Y, Jiang G Y, Yu M, et al. 3D visual discomfort predictor based on subjective perceived-constraint sparse representation in a 3D display system[J]. *Future Generation Computer Systems*, 2018, 83: 85-94.